

**Engagement Opportunities in NASA STEM 2022 (EONS-2022)**  
**NASA Research Announcement (NRA)**  
**MUREP PBI/HBCU Data Science Equity, Access and Priority in Research and Education**  
**(MUREP DEAP)**  
**Number: NNH22ZAO001N-MUREPDEAP**

**Title:** DEAP Institute in Research and Education for Science Translation via Low-Resource Neural Machine Translation

**Institution:** Prairie View A & M University

**City/State:** Prairie View, Texas

**PI:** Lijun Qian

**FY:** 2023

**Summary:** The goal of this project aims at establishing a DEAP institute (consortium) and building an AI-based system that can share interactive, instantaneous, and user relevant Earth science information to allow NASA science to be more discoverable and more accessible to a broad audience. The research objective of this project is to map water related questions from non-scientists into database queries. It can be defined as a sequence-to-sequence mapping problem, where the source sequence includes layman's terms on the water related questions while the target sequence will be most relevant database queries. We plan to complete this task by developing a novel mapping method based on low-resource neural machine translation (NMT). NMT has emerged to significantly promote machine translation with end-to-end models that automatically translate a source language to a target language through mapping source sentences to target sentences. Because large amount of data like pairs of annotated question and queries may not be available, low-resource NMT techniques will be developed for this project. Specifically, we plan to build a semi-supervised NMT based on our preliminary works on semi-supervised learning and natural language processing. It will enhance the mapping performance by leveraging large amounts of unlabeled data. The deliverables include 1) Docker image of semi-supervised NMT based question-query mapping; 2) online website that is integrated into the ongoing Earth Information System that allows us to demonstrate the proposed solution to stakeholders at NASA and the community. The proposed project will greatly improve the research and educational capacity at multiple HBCU/MSIs, enable broad participation of underrepresented minorities in advanced research, build strong partnership with NASA, and provide the much-needed diversity in the next-generation workforce for the nation.